# Students Today, Teachers Tomorrow? The Rise of Affordable Private Schools

## Preliminary Draft: Please Do Not Cite

Tahir Andrabi (Pomona College) Jishnu Das (World Bank)
Asim Ijaz Khwaja (Harvard University)*

November 2005

**Abstract**

The debate on achieving universal primary education has largely ignored the role of the private sector. Yet private schools play an increasingly significant role in several low-income countries by offering affordable schooling options for even the rural poor. This paper uses data from Pakistan — a country that has experienced a mushrooming of mainstream, for-profit, and affordable private schools — to examine where such schools arise. We identify a large externality generated by the public sector: Private schools are set up in villages where there are pre-existing public girls' high schools. Instrumental variable estimates suggest that a girls' high school increases the likelihood of a private school in the village by 35 percentage points. In contrast, there is little or no relationship between private school existence and pre-existing girls' primary or boys' primary and high schools. The data support a supply side explanation: in an environment where female mobility is low due to cultural restrictions, women receive significantly lower wages in the labor market. Private schools locate in villages with a greater supply of local high-school educated women and pay them low wages. Our findings show that the private school wage-bill is indeed lower in villages with pre-existing girls' high schools. These findings bring together three related concepts—the inter-generational externality of public schools on the existence of private schools, the ability of the private sector to use cultural labor market restrictions to it's advantage and the prominent role of women as teachers.

*Pomona College, Deverlopment Research group, World Bank, and Kennedy School of Government, Harvard University. Email: tandrabi@pomona.edu; jdas1@worldbank.org; akhwaja@ksg.harvard.edu. This paper was partly funded through grants from the PSIA and KCP trust-funds and the South Asia Human Development Group at the World Bank. We thank Ghazala Mansuri and Tara Vishwanath for early discussions. We are specially grateful to Tristan Zajonc for many, many hours spent in matching data at the village level. Assistance from the Project Monitoring and Implementation Unit in Lahore is also acknowledged. All errors are our own.

# I Introduction

The two Millennium Development Goals (MDGs) related to education call for universal primary education by 2015 and eliminating gender disparities in primary and secondary education. By 2005, most countries had already fallen well behind the necessary targets to meet these goals. While this has led to renewed calls for greater public investment in schools through school construction, teacher training programs and cash transfers, the role of private schools has been notably absent from the debate. This is particularly surprising given that the private sector is a significant an increasing presence in primary education in several low-income countries.

This paper looks at the private schooling sector in Pakistan, taking advantage of a unique census of private schools. In 2001, one in three children enrolled in a primary school in Pakistan was in a private school, with an increasing share from rural areas and from middle-class and poorer families. These mainstream and unaided private schools, funded almost entirely through low school-fees, exploded during the nineties, with the share of private schooling in primary enrollment doubling. Pakistan's experience is not unique—recent findings suggest that India and Bangladesh are going through similar changes. However, the growth in private schools has not been uniform and exploiting this variation provides insights into the working and limits of the private sector.

We focus on a specific question — what determines where (affordable) private schools arise? In a context where most low-income countries' public sectors are seriously over-stretched, understanding whether the private sector can become an instrument for mass education is clear important. Our examination also speaks to the important role of the public sector, of women and supply side constraints. We identify a large positive externality from pre-existing girls' public high schools to private schools. We argue that this externality arises through limited labor markets for educated women. Private schools are able to operate widely and profitably despite charging low fees, by taking advantage of a low-cost teaching resource — locally educated women willing to work at low wages due to restricted labor market opportunities.

We argue that the initial conditions for the increase in private schooling in Pakistan were laid during the 1980s, with a systematic policy of public high school expansion for girls.[1] Private

[1] By "high" school we mean schools that offer classes above primary, i.e., from Grades 6 to 12 in addition to classes at the primary level.

schools were more likely to locate in villages that received a girl's high school during this period of school expansion; similar results do not extend to girl's primary schools or boy's schools (primary or high). The numbers are striking: among villages that did not have any girl's school in 1981, by 2001 there were private schools in only 11.5 percent of those that still did not have a girl's school, 12.5 in those that received a girl's primary school and 30.8 in those that received a girl's high school.

This positive association between private schools and girl's high schools could reflect non-random placement patterns. If girl's high schools were constructed in villages with high aggregate demand for education, estimates reflect differences in demand rather than the causal impact of girl's high schools on private schooling. Similarly, if they were built in villages with systematically lower wages, a positive correlation could reflect aggregate differences in labor market conditions whereby teacher salaries in private schools were lower than in random villages.

We address non-random selection in a number of ways. We match village-level data from 1981 and 1998 and show that apart from village population, there were no significant baseline differences in village-level characteristics for villages that did and did not receive a girl's high school. These data also allow us to remove unobserved time-invariant characteristics through a first-difference specification. However, these methods do not address the issue that the placement of girl's high schools could be correlated with time-varying village characteristics.

Our preferred specification addresses concerns arising from time-varying omitted variables through an instrumental variables approach. We use stipulated rules for the placement of girl's high schools to construct an "intention-to-treat" estimator. Government rules specified a number of conditions to determine whether a village was eligible, which we argue can be modelled as a "minimum population condition" and a "population rank condition". In particular, villages satisfy an eligibility condition if (a) their population was at least 500 and (b) their population was the largest within a geographically contiguous administrative unit of 3-4 villages, called a patwar-circle.

Together, these conditions generate an intent-to-treat variable that is highly correlated with the actual placement of a girl's high school—villages in the eligible group were 60 percent more likely to receive a girl's high school compared to those that were ineligible. Further, non-linearities in the eligibility requirement (two villages with same population may be eligible or not depending on their population rank within the administrative unit) allow us to control for linear (and polynomial) effects of village and neighboring villages' populations that might have independent effects on private school placement. Under the assumption that private school placement is not determined in the same non-linear fashion as the eligibility requirements, the estimated coefficient from this instrumental variables approach is consistent and unbiased.

From a base estimate of 10 percentage points (75 percent) in the cross-sectional data, the impact of girl's high schools on the existence of a private school increases to 15 percent once we control for time-invariant village-level characteristics and to 35 percent with the instrumental variables specification. This increase suggests that time-varying omitted variables were indeed important in the placement of girl's high schools. We postpone a discussion of why this may be so till later.

What are the channels through which girl's high schools lead to an increase in the likelihood of private school existence in the village? In the words of a local entrepreneur:

*"The big problem"*, he said*, "is teachers. In most villages, I can set up a private school, but who will teach? All the men are working and if I pay them what they want, I will never make a profit. I cannot get women from other villages—who will provide the transport for them if it gets dark? How will she be able to work in another village if she is married? The only way we can work is if there are girls who can teach in the village—that is why, I ask if there is an educated girl who can teach. I can pay them Rs.800 ($14) a month and run the school. Otherwise it is not possible."*

Taken at face-value, our entrepreneur's travails suggest that one channel through which girl's high schools led to the expansion of private schools was by creating teachers. Of course, this need not be the only one—educated women also play an important role in stimulating demand both for their own children and those around them. In addition, girl's high schools could also increase awareness or change preferences towards education in the village and these may be different from what girl's primary or boy's schools can do.

Nevertheless, the entrepreneur's suggested hypothesis is not easily dismissed. In 1981, there were 5 literate women in the median village in Punjab—the largest and most dynamic province in the country. High-school education was a rarity: 60 percent of villages in the province had 3 or less such women and 34 percent had none. While hiring men is clearly an option, data on private school teachers suggests that this would not come cheap—wage differentials between men and women controlling for education, age and experience suggest a 25-30 percent discount for female teachers. In addition since private schools are all coeducational, the presence of female teachers could be instrumental in getting more girls into school. Interestingly, an (early) innovative randomized experiment that tried to set up private schools in rural areas failed precisely because teachers could not be found (Alderman 2003).

In an independent survey, we do indeed find that most private school teachers are females who were educated in the same village as the one they are teaching in (Andrabi and others 2005b). With few employment opportunities outside teaching (both due to the lack of options and social constraints on what women can and cannot do), private entrepreneurs could act as monopsonists in the villages where they set up; conversely they would be severely limited if they tried to locate in villages without high school educated women.

Our results provide further support for this "women-as-teachers" channel. Recall, that private school existence is only affected by pre-existing girls' high schools - girls' primary or boys' high/primary schools have little or no effect. For this to be explained through a demand channel one would have to assume that higher educated women stimulate demand for education in the village much more than equally educated men or women/men with lesser (primary) education. More directly, our results show that not only do pre-existing girls' high schools more than double the fraction of highly educated local women, but that private schools are more likely to exist when there are more high-school educated women in the village, whereas similarly educated village men have no effect.

However, these predictions are on the "quantity" margin and more nuanced demand based channels could also explain the observed location patterns of private schools. A more compelling test is on the price/wage-bill margin. If the supply-side channel is more important, one would expect the wage-bill for setting up a private school is lower in villages with pre-existing public girls' high schools - the demand side channel would suggest the opposite. Examining data on private school wage-bills, shows that this wage bill is indeed 20 percent lower in villages with a girls' high school.

While there is burgeoning interest in the relative performance of children in public and private schools in the US (Hoxby 1994, Altonji 2005) and in low-income countries (Jimenez and others 1999, Kingdon 1996 and Tooley 2005), we know less about where such schools setup in the first place. In the US there is some evidence that the religious affiliation of communities plays an important role in where private schools locate (Hoxby 1994, Downes and Greenstein 1996) and a recent paper from India confirms patterns similar to those documented here, whereby private schools are more likely to locate in communities with public schools (Tilak and Sudarshan 2001). To our knowledge this is the first paper that establishes a causal cross-generational link between public school construction and the future development of private schools. In contrast to findings that public schools deter private school entry, our results suggest that the supply-side effects of producing educated teachers outweighs the loss in demand from existing public schools.

In related research on private schooling we show, as highlighted above, not only are private schools in Pakistan very affordable and cost-effective with a fee-to-GDP ratio of 4 percent (compared to 14 percent for the US), but that children studying in private schools are learning a lot more than in public schools in the same village—the difference between children in private and public schools varies from a half to one one standard-deviation in test-scores depending on the subject (Andrabi, Das and Khwaja 2005a,b). These results are virtually unchanged with additional controls for household wealth, parental education, distance to school and the number of siblings. Previous research documents very similar patterns with higher quality-to-cost ratios in private compared to government schools for a number of other low-income countries (Jimenez and others 1999, Kingdon 1996 and Tooley 2005).

In addition, the Pakistani data also show that gender ratios are more balance in private schools. Private Schools offer mainstream and english-medium education and there is suggestive evidence that they raise overall enrollment levels by offering parents greater and closer schooling options. Together these findings emphasize the importance of understanding the role private schools can and are playing, particularly in low-income countries.

The paper is structured as follows: section II is a guide to the educational background and institutions in Pakistan. Section III presents some patterns that illustrate our main findings. Section IV discusses the data and Section V details our methodology. Sections VI presents the results. Section VII concludes.

## II Background and Institutions

Pakistan has a population of 132 million people living in four main provinces—Punjab, Sindh, North Western Frontier Province and Balochistan. They, along with Islamabad the federal capital, comprise about 97 percent of the country's population. Punjab, the focus of our study, is the largest province with 56 percent of the population, a majority of whom are in rural areas. Education provision is managed by provincial governments but there are recent changes at the local government level that aim to devolve such provision to the lowest tier.

Both the literacy rate and primary school enrollments are low, even compared to countries in Pakistan's own income range and to others in South Asia. Pakistan's adult literacy rate of 43 percent compares to 54 percent for the South Asia average. The gender gap in educational enrollment is large as are differences between rural and urban areas and the rich and poor. The gross enrollment ratio for the top expenditure decile is twice as high as that for the lowest decile. Increasing the stock of human capital is then clearly a priority, especially for the poor, and especially for girls.

The extent to which supply-side interventions, such as the construction of schools, can help in increasing human capital is constrained by the nature of the labor market. While returns to education in Pakistan appear to be fairly high at 15.2 percent (Psacharopolous 2002) and higher for women compared to men (World Bank 2005), labor force participation for women is low. Labor force participation was less than 5 percent for women in rural areas according to a household survey fielded in 2001 (Pakistan Integrated Household Survey) and of those employed in paid jobs, ?? percent were working as teachers. Moreover, wages for women were 30 percent lower than for men after controlling for educational qualifications and experience (World Bank 2005). Low mobility for women appears to play a role—a recent study by the World Bank documents that safety concerns and a strong patriachal society restrict the ability of women to find work outside the village they live in.

The government's position on the role of private schools has oscillated during the last 30 years. Private Schools were nationalized in 1972 amidst a mass government program of nationalization of all industry, but in 1979 the policy was reversed. Private schools were allowed to open, and schools taken over by the government were gradually returned to the original owners. However, there was no systematic policy towards such schools or any subsidy in the form of grants (to parents or directly to schools) unlike in India or Bangladesh (where there are a large number of private *aided* schools—the Indian equivalent to the Pakistani private schools are the Private Unaided Schools). Prior to nationalization, private schools catered primarily to a niche market, restricted to large cities. The market was dominated by missionary run private schools (or local schools imitating the missionary model), mainly used by the elite.

Within this milieu, the government in consort with international donors tried to increase investment in human capital through a series of "Social Action Programs" or SAPs. The SAP initiated in 1980 called for large investments in the provision of education. School construction was a large part of the education component, and specific guidelines were set for villages where schools should (and should not) be built.

Figure I shows that this program had an effect, at least on school construction. The figure plots the number of years a school has existed for all villages in Punjab and for four types of schools—primary and high for boys and girls. The kernel density shows a substantial increase in construction around the years of the SAP III program; in the case of girls, this is the largest construction period since 1930. While most girl's high schools seem to have built between 1960 and 1965, there is again a definite hump around the SAP III years; in the case of boy's high schools, there is no such increase—most such schools appear to have been around since 1930.

**Why do private schools matter?**

Since this paper focuses on understanding what factors contribute to the creation of private schools, it is worth emphasizing why this question is important, both in general and in the Pakistani context.

While the 2000 Millennium Development Goals debate on achieving universal primary education and eliminating gender disparities in education has focused mostly on public schooling and cash transfers, there is now a beginning realization that private schooling can and is playing an important role in achieving these goals. Private schools play an important role in primary enrollment not just in Pakistan with over a third enrolled in the private sector in 2000 - but also in India and Bangladesh in South Asia and other countries such as Zimbabwe in Sub-Saharan Africa and Lebanon and the gulf-states in the Middle East.

Part of the reason private schooling has not received as much attention is that private schools are often thought to be expensive and urban, catering only to the rich and that too at the expense of the public sector. Hence their impact on raising overall enrollment, particularly in rural areas, was thought to be limited. However, in related work using detailed data from Pakistan (Andrabi et. al. 2005a-c) we show that these assumptions are not correct on several important counts.

First, private schools are surprisingly affordable and are an increasingly rural phenomena (Andrabi et. al. 2005b). A typical private school in a rural village in Pakistan charges Rs.1000 ($18) per year, which represents 4 percent of the GDP per capita for the country. In contrast, private schools (elementary and secondary) in the US charged $3524 in 1991. At 14 percent of GDP per capita, the relative cost of private schooling is almost 3.5 times as high in the US compared to Pakistan. While the majority of private schools setup in Pakistan before 1990 were urban, since then, there has been a steady increase in the rural/urban ratio with equal numbers setup in rural and urban areas in recent years. Perhaps as a result of their affordability, private schools are increasingly accessed by even those from the lowest income decile.

Second, private schools, while offering comparable if not better facilities (per student class-rooms, black-boards, sitting arrangements etc.), incur half as much expenditures per child compared to public schools - Rs. 1,012 per child for the median private school as compared to Rs. 2,039 for the median public one (Andrabi et. al. 2005a). These differences remain just as large even after controlling for parental/village wealth and education. Given that teaching costs form the bulk of schooling costs, it is not surprising that this saving stems primarily from lower teacher salaries in the private sector. These lower teaching costs will form an important part of the analysis in this paper and we will revisit this in more detail below. If private schools offer comparable quality to public ones, examining private schools may provide a means to lower the overall cost of education provision without compromising on quality.

This brings us to the third important point: Private schools seem to offer better educational quality than public ones. In tests administered by our team to over 12,000 students finishing class 3 in over 800 schools in 110 randomly selected villages with a private school in rural Punjab, we find that private schools students outperform public school students (Andrabi et. al. 2005a). The differences are large with private school students typically scoring a standard deviation higher in mathematics, Urdu (the national language) and English, with the largest differences in English. The differences remain as large even after controlling for child, household, and village attributes.

Finally, to what extent do private schools contribute to overall enrollment rates and education quality? These are empirically difficult questions. However, while private schools are indeed a substitute to public schooling, the limited evidence suggests that they may play a beneficial aggregate role. First, to the extent that public schools in developing economies are over-crowded, private schools can lessen the burden on the public sector and may raise educational quality. Second, by offering a richer and potentially higher quality choice set, private schools may raise overall enrollment levels. While not the focus of this paper, our preliminary analysis suggests that overall (public and private) enrollment rates in villages that have a private school are higher than those without, and that these results are not driven by differences across villages (in demand for education etc.). In related work we also show that the single largest factor that affects the enrollment decision is distance to school, with even larger effects for girls (Andrabi et. al. 2005a,c). Since the public sector school placement guidelines discourage more than one public school in a village/settlement, private schools may raise overall enrollment simply by offering closer schooling choices.

A point worth noting in the Pakistani context is that these schools are not religious schools or madrassas. The vast majority of these schools are co-educational, English medium schools which offer secular education. A recent study by the authors (Andrabi, et. al. 2005c) shows that, contrary to popular views, religious schools play a much smaller role in Pakistan, comprising a less than one percent share and an even lower share in settlements with a private school.

## III Some Illustrative Patterns

Three figures and a table outline the basic patterns of interest for this paper. Figure II shows the date of construction for every private school based on a private schooling census carried out in 2001. A quarter of the private schools in 2000 were built in the previous year, and there is an exponential decline as we move further back in time. During the early half of the nineties, most schools were in urban areas. However, there was a shift around 1995 with an increase in the share of rural villages. This pattern continues till 2000 (the last year for which we have data) when approximately half of all schools were set up in rural areas. Although these data confound school construction and survival, comparisons with the number of schools reported in a 1985 study (Jimenez and Tan 1985) suggest that most of the increase is real—correcting for the survival rate (around 5 percent of all schools die every year) still implies a phenomenal growth rate during the nineties. An important part of our analysis will be to try and understand where such private schools arise.

Figure III shows the relationship between the existence of a private school and various types of government schools. We regress the existence of a private school on the number of years that the village has had a primary or high school (both boys and girls) using a probit specification. The figure plots the predicted probability of a private school against exposure to a public school; these probabilities are then to be understood as the marginal effect of exposure to each type of public school, controlling for other public schools in the village.

Boy's primary schools appear to do nothing with an almost flat relationship over the 20-year horizon. Girl's primary schools and boy's high schools do marginally better and taken at face-value, the regression implies a 2-3 percentage point increase in the probability of a private school with 10 years of exposure. The role of girl's high schools stands out. There is a private school in one-fifth of all villages with a girl's high school, or a 7-8 percentage point increase for every additional 10 years of exposure. The marginal impact of a girl's high school on private school existence is large and significant.

Figure IV is another cut at the data showing the predicted relationship between private school existence and the percentage of men and women with high school education (8 or more years of schooling) in every village. Increasing the percentage of women with high school education in a village increases the probability of a private school from 10 percent (0 percent educated) to almost 100 percent (100 percent educated). Educated males again play an attenuated role—at the mean of the sample for female literacy, increasing high school males from 0 to 100 percent increases the probability of private school existence from 7 to 32 percent.

Both these figures suggest that women matter—girls' high schools (henceforth GHS) have a qualitatively different role in the setting up of private schools as do educated women in the village. In addition, labor market conditions may be important. As part of a longer-term investigation (the LEAPS project), we interviewed 5000 teachers in government and private schools from a random set of villages drawn from a list frame of villages with at least one private school. Comparing wages for women and men in private schools show that the former are paid 25-30% lower conditional on location, educational qualification and age (Andrabi and others 2005b).

# IV Data

The above suggests that pre-existing GHS lead to private schools being setup by lowering the cost of teachers. To examine the degree to which these correlations are robust and causal, we need data on where private schools are located today matched to where (different types) of public schools are located. In addition contemporaneous data on village-level attributes would allow us to control for (observed) confounding factors correlated to the existence of private schools. Finally, data from the pre-construction (of public schools) period can constitute a "baseline", which allows for more flexible specifications and help us disentangle correlation from causality.

We therefore employ four different data sources: (a) a complete census of private schools; (b) administrative data on the location and date of construction of public schools and (c) data on village-level demographics and educational profiles from the 2001 and 1981 population censuses. The Federal Bureau of Statistics undertook a census of all private schools in 2001 and data on public school construction is available from provincial Educational Management Information Systems (EMIS). Further, population censuses carried out in 2001 and 1981 (two years after the denationalization of private schools) provide both contemporaneous and baseline data on village-level characteristics.

An important step was matching these dataset at the village level. Given the high level of data fragmentation (these data are collected by different institutions and do not support a common coding scheme) and variations in the spellings for village names some choices had to be made. By relying on phonetic matching algorithms and a manual post-match, we matched the public school (EMIS) data to both the 1981 and 2001 censuses. Given the demanding nature of the task we choose to do the matching only for Punjab province and only for rural areas. Our matched sample therefore, is 18,119 villages out of 25,941 in 2001 in Punjab province with a population of 42.3 million in 1998, representing 84 percent of the total rural population of the province.

From this matched sample, we employ two restrictions to generate our final sample. First, our empirical strategy relies on the availability of village-level baseline data *prior* to the construction of a public school in the village—for villages with pre-existing public schools, we cannot discern whether differences in the baseline data arise from selection into villages or the exposure to a public school. Thus, we first restrict our sample to those villages that had *not* received a girls high or primary school by 1981.

We are also concerned that the presence of pre-existing girls' high schools in *neighboring* villages could bias our results. Pre-existing girls' high schools in nearby villages could affect the demand for education even if the concerned village does not have a school in the baseline data. Our final sample therefore consists of all geographical units, known as patwar-circles, within which *no* villages had a girls' high school prior to 1981.[2] In the absence of spatial data on the location of villages, patwar-circles, which are typically a group of 2-4 geographically contiguous villages, are a plausible measure of the local availability of schools. Preliminary findings indicate that our results are similar across the restricted sample and the full sample; we are in the process of investigating whether the villages we could match across the datasets differ in a systematic manner from the villages that we could not (and the implications for our results).

[2] We are less worried about girls' primary schools in neighboring village affecting village demand, since there is considerable evidence that younger children do not travel outside their village to go to school (Alderman, Jacoby and Mansuri, Andrabi and others 2005b).

Table I presents some summary statistics. Mirroring the construction of public schools (Figure 1), there has been a boys' public school in the average village for 27 years, a boys' high school for 6 years and a girls' primary school for 7 years. In contrast, exposure to a girls' high school is 0.7 years and this reflects the large number of villages without such a school. There is a private school in one out of every 7 villages and, continuing with the low availability of educated women, in the average village there are only 13 women (out of a population of 1830) who report secondary or higher education in 1998.

# V Methodology and Empirical Framework

A simple framework outlines the private entrepreneur's problem, focusing on the role of the public sector and the econometric and interpretational issues in identifying the impact of a GHS. Consider an entrepreneur's decision to open a private school in village $i$. She will open a school if it yields a positive net present value.[3] We consider a static version of the problem so that a private school is opened in village $i$ if it yields positive net returns in that period. This is analogous to an assumption that there are small fixed costs in operating a private school and rules out dynamic considerations. The per-period net revenue is then the difference between total revenues and total costs. For private schools, 98.4 percent of total revenues come from school fees and 89 percent of the costs are salary costs for teachers (Andrabi and others 2005c). Thus, we write net return as:

$$NetReturn_i = Fee_i * N_i - Wage_i * T_i \tag{1}$$

where $Fee_i$ is the average private school fee for a single student, $Wage_i$ is the average private school teacher's salary and $N_i$ and $T_i$ are the number of students enrolled and teachers employed in the private school. An important feature of our analysis is that limited female mobility across villages results in segmented labor markets and therefore a cost advantage for entrepreneurs in one village over another. We thus allow both the school fees charged and teacher wages to vary across villages.

[3] This assumes that there is no shortage of entreprenuers (otherwise not every positive NPV project will be undertaken). While we can incorporate such shortages, doing so will not change the qualitative results.

We further posit that the total revenue a private school earns (the demand side) is a function of village demographics (increasing in population or adult literacy) and alternative schooling options (number of pre-existing public schools or the travel costs from the village to neighboring villages with schools). Wage costs in a private school (the supply side) also depend on village characteristics if labor markets are not perfectly integrated across villages. Thus wages are likely to fall with greater availability of local potential teachers, which in turn may be affected by pre-existing public schools or lower travel costs. A reduced form expression for net return is then:

$$NetReturn_i = \alpha + \beta_1 GHS_i + \gamma_1 GHS_i \; + \beta' X_i^D + \; \gamma' X_i^S + \varepsilon_i \qquad (2)$$

where $X_i^D$ and $X_i^S$ are village demographics and characteristics that respectively affect the demand for private schooling and the costs of running such schools. Our variable of interest, $GHS_i$, captures the presence of a GHS. It appears twice in equation(2) to emphasize that pre-existing public schools affect both the demand for private education (negatively by acting as a substitute and positively by creating a more educated populace) and the cost of setting up such schools (by affecting the local supply of potential teachers).

Since we do not observe the net return a private school earns, we treat net return in equation(2) as a latent variable in a probability model, so that $Prob(PrivateSchoolExists) = Prob(NetReturn_i > 0)$. If the error term is uncorrelated to the existence of a public school, OLS estimation of equation(2) recovers the joint effect $(\beta_1 + \gamma_1)$ of higher demand and lower costs from GHS construction on the existence of a private school. In Section VI, we argue that the impact of GHS construction on private school existence is consistent with a supply side channel; this tilts us more towards an explanation based on the coefficient $\gamma_1$.

This admittedly simple framework highlights the main empirical issues. The primary concern in estimating the impact of GHS on private school existence is that villages that received GHS may be different from those that did not in a manner that directly affects net return, or, $cov(\varepsilon_i, GHS_i) \neq 0$. This can happen either if villages that received GHS also had a higher demand for education or if they had systematically lower labor costs.[4] To the extent that such

[4] These two biases typically move in opposite directions, particularly if we assume (as is reasonable) that the demand for education is correlated to future returns in the labor market.

differences are not observed and controlled for, the OLS estimate of $(\beta_1 + \gamma_1)$ is biased and inconsistent.

We deal with this in a variety of ways. Our first specifications use the two periods of data to control for time-invariant omitted variables. One alternative is to regress the existence of private school on current and lagged village level characteristics, the setting up of public schools, and geographical dummies, $D_j$, at the level of the patwar-circle.

$$\Pr iSchool_i = \alpha + (\beta_1 + \gamma_1)GHS_i \ + \beta' X_i^D + \ \gamma' X_i^S + \beta' X_{i(lag)}^D + \ \gamma' X_{i(lag)}^S + \sum_{j=1}^{M} \delta_j D_j + \varepsilon_i \quad (3)$$

A second option is to estimate a first-differenced specification at the level of the village.

$$\Delta \Pr iSchool_i = \alpha + (\beta_1 + \gamma_1)\Delta GHS_i \ + \beta' \Delta X_i^D + \ \gamma' \Delta X_i^S + \varepsilon_i \quad (4)$$

These two specifications differ in what they are able to account for. In equation(3), we allow for differential growth rates in net return across patwar-circles and control for village-level omitted variables through the inclusion of lagged demographic characteristics. The coefficient on $GHS_i$ is unbiased if there are no unobserved village-level omitted variables. In equation(4), we difference-out all time-invariant village-level variables but, with two observations per village, cannot control for differential village-specific growth rates (although we can and do include area/PC-wide time trends). Finally, we can also implement the non-parametric equivalent of equation(3) through first-differenced propensity score matching techniques. That is, we can compare the change in the number of private schools across matched villages that did and did not receive a public school, where the matching is implemented on the baseline data.

While these strategies address biases arising from time-invariant omitted village-level characteristics, unobserved time-*varying* village attributes could also play a role. Even if two villages are identical to begin with, one village may experience a change that increases the likelihood of receiving a government school and simultaneously raises the returns to a private school. For example, the demand for education may increase in a village when a new road creates better job opportunities. This change leads to greater demand for a public school and higher returns to private schools. Although we partly address this by allowing for common trends at the level of patwar-circles, as in equation(3), our particular setting and the program through which the public school construction was undertaken provides a promising instrumentation strategy. This strategy uses official rules for public school placement to construct an intent-to-treat estimator.

### A. A Rule-Based Instrumentation Strategy

Our instrumentation strategy follows Campbell [1969] and Angrist and Lavy [1999]. The strategy exploits the fact that the regressor of interest, in our case the construction of a government school, is partly based on a deterministic function of a known covariate; in our case, village population. If this deterministic function is non-linear and non-monotonic, it can be used as an instrument while directly controlling for linear and polynomial functions of the underlying covariate itself.

School construction in our sample was a direct consequence of the Pakistan Social Action Program in 1980. While the majority of girl's schools setup during this time were at the primary level, several villages also received GHS. These secondary schools were not add-on's to an existing primary school. Under the SAP, a certain number of GHS had to be established, and the government could do one of three things: (a) they could either not set up any school, (b) they could construct a girl's primary school or, (c) they could construct a GHS that also included a primary section. In this sense, the construction of a GHS really meant that the building and sanctioning of teachers followed both the protocols for a primary and secondary school—there would be more classrooms and more teachers, although in the initial years, the number of children in the secondary classes was likely to be very low. Reflecting this design, out of the 328 villages in our sample that received a GHS between 1981 and 2001, only 31 had a pre-existing girls' primary school; in all the rest, the secondary and primary sections of the school were constructed simultaneously.

Of interest to us is that there were specific guidelines for where these schools could (and could not) be built. For GHS the official yardstick for the opening of new schools was that (i) the population of the village be no less than 500 (ii) there should be no government GHS within a 10 km radius and (iii) the village would have to provide 16 kanals of land.

In order to exploit the nonlinearity in these guidelines we construct a binary assignment rule,

$Rule_i$, that exploits the first two guidelines. Within our sample restriction, $Rule_i$ is assigned a value 1 if (and only if) the village's population in 1981 was greater than 500 *and* it was the largest village (in terms of 1981 population) amongst nearby villages. The latter captures the radius criteria: if a village is not the largest village amongst its neighbors, it is likely that its neighbor would receive a public school first given the stated preference for population in the construction of schools. Provided this school is near enough, the village will be less likely to receive its own public school.[5]

Ideally we would have liked to use actual distances between villages. In the absence of geographical data, we proxy this by including all the villages that are in the same "patwar circle" ($PC$) as approximating the radius conditions of the rule. In terms of actual land area, this is a reasonable approximation—given the size of the province and the number of patwar circles, a back of the envelope calculation establishes that a school within every patwar circle would (roughly) satisfy the radius requirements of the rule. Our instrument then takes a particularly simple form: a village is in the "intended-to-treat" group if it had a population of at least 500 in 1981 and it was ranked first in 1981 population within the patwar-circle. Formally:

$$Rule_i = \begin{array}{l} 0 \; if \; Population_i^{81} < 500 \\ 1 \; if \; Population_i^{81} \geq 500 \; AND \; Population_i^{81} = \max_{j \in PC_i} (Population_j^{81}) \\ 0 \; if \; Population_i^{81} \geq 500 \; AND \; Population_i^{81} < \max_{j \in PC_i} (Population_j^{81}) \end{array}$$

This rule is non-linear (it jumps from 0 to 1 at a fixed population threshold), discontinuous, and non-monotonic—it drops to 0 for larger villages when there is an even larger neighboring village within the patwar circle. In using this rule as an instrument, we can explicitly control for continuous functions of a village and its neighbors' populations since these covariates may also have a direct impact on our outcome of interest, the existence of a private school.

To clarify the identifying assumptions, consider how our instrument relates to 1981 popu-

[5] Another alternative is to use the radius-rule directly and assign $rule_i = 0$ if there is a village in the patwar-circle that has a GHS. This is problematic since we are worried about the endogenous placement of GHS in the first place.

lation and how population in turn affects our outcome of interest. Figure V illustrates how the binary instrument covaries with a village's population in 1981. While all villages below 500 are assigned a value 0, there are a substantial number of villages above this threshold that are not in the intended-to-treat group—these are villages that are not top-ranked within their patwar circle. This is reassuring since it indicates that the instrument varies even within villages of comparable size.

Figure VI shows the non-parametric relationship between village population and the probability that it receives a private school. The predicted relationship is much smoother than between population and our instrument. Since village population clearly has a direct effect on the likelihood of having a private school, once we condition on smooth functions (polynomials) of population, the variation we exploit in the instrument is precisely the remaining nonlinearity (the jump at 500 and sudden drops at higher populations).

Put another way, our strategy compares two villages with the *same* population but with different population ranks within the patwar circle. The first stage then reflects the extent to which population ranks affect the probability of receiving a girl's public school and the reduced form is the extent to which population ranks affect the likelihood of having a private school. The exclusion restriction is that, for two villages with the same population, the relative population rank of the village within the patwar circle does not affect the probability of having a private school, except through GHS construction.

We can think of two reasons why this exclusion restriction may fail. First, if private school entrepreneurs search for the highest returns and choose among villages within a patwar circle (but not a broader area), they will likely choose the village with the highest population. However, this relies on two implicit assumptions: (a) that there is a shortage of (local) entrepreneurs, so that even in villages where the net present value of doing so is positive, a school is not set up and (b) that private entrepreneurs need not be resident in the village where they set up the school. Neither assumption is particularly plausible. There is no reason to suppose a shortage of entrepreneurs particularly since there are few fixed costs involved. Similarly, the majority of private schools in Punjab province are small, family-owned units located in the entrepreneur's residence. This allows the entrepreneur to economize on rents and to develop strong relationships of trust with the community. The latter is critical—given safety concerns and mobility constraints, parents are unwilling to send girls to schools or let unmarried girls teach in schools where the head-teacher is not known to the community.

Our exclusion restriction could also fail if the government used the same rules for allocating *other* investments that may affect the return to private schooling in the village. We deal with this in two ways. First, in all specifications, we control for the presence of all types of government schools in addition to GHS, which could affect the probability of private school existence.[6] Second, we include regressors that are plausible measures of other government investment, such as electrification and water-supply. We find no evidence that the rank of the village affected the level of government investment, either in means-comparisons or in regression specifications. Indeed, other government investments such as water and electricity are slightly *lower* in higher population rank villages, although the coefficients are small and insignificant (results available with authors).

A final econometric issue is that both our treatment variable (the construction of a public school) and our dependent variable (the existence of a private school) are binary. Early work in the choice of specifications suggests that linear instrumental variables estimates are not very different from more structured models in such settings (Angrist 1991). However, in addition to the limited dependent variables, our treatment probabilities are also low. At low treatment probabilities the choice of specification will make a difference. Linear specifications assume a uniform error term while non-linear specifications impose a normal or logistic distribution on the error. When treatment probabilities are neither high nor low, the cumulative distributions of the uniform and normal (roughly) match up, yielding very similar results. With low treatment probabilities though, the cumulative distributions differ considerably and estimates from linear IVs could give very different results from alternative, non-linear specifications.

Following the literature, we propose two sets of estimates based on a linear and a bivariate probit specification. The bivariate specification requires the assumption of joint normality of

[6] A potential issue here is that our instrument may also predict the construction of other (than girls high) types of public schools. However, once we condition on a village and its PC's maximum population in the restricted sample the rule predicts post-1981 construction only for GHS.

the error term in the determination of public and private schools and can be formally stated as

$$Pub_i = 1(X^{'}\beta + u > 0) \tag{5}$$

$$\Pr i_i = 1(X^{'}\gamma + aPub_i + \delta Rule_i + \varepsilon > 0)$$

$$\begin{bmatrix} u \\ \varepsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \tag{6}$$

We are currently in the process of implementing semi-parametric versions that impose less structure on the empirical model (Abadie 2005), and will incorporate the results in the next version.

## B. Isolating the Supply Side

The proposed instrumental variables strategy isolates the causal impact of public school creation on private school existence. Separating out supply-side from demand-side channels is harder. Pre-existing public schools simultaneously create more educated parents (increasing demand) and the potential supply of teachers (increasing supply). Even an exogenous increase in the number of unmarried and educated adult women in a village (say, through a helicopter drop?) does not allow us to easily separate demand from supply channels: such women, even if they don't have children of their own, could influence the demand for education among their immediate family and neighbors. Thus, while our results do identify the causal impact of pre-existing public schools on the existence of private schools, relying purely on the quantity margin i.e. the existence of private schools - we would be unable to further isolate the precise channel with existing data.[7] However, our data also allows us to test predictions on the price margin. If GHS result in a lower wage bill for private schools, we should see that villages with a GHS have lower costs in hiring women teachers. We will directly test for this as well.

[7] One possibility is to use the variation in the timing of the public school construction. If supply-side stories are correct, we would expect a private school to emerge 5-8 years after the construction of the girl's high school. Unfortunately, that data are too limited to exploit such variation. We require villages with both private schools and girl's high schools. Since only 324 villages received a girl's high school and of these 30 percent had a private school, we are unable to identify any discontinuities off the 100 or so villages that have both. Another possibility is to check whether there is a difference in the existence of a private school based on years of exposure to a GHS. Here we do find some evidence that less than 5 years of exposure has no effect on the likelihood of private school existence. In particular, private schools exist in 18 percent of villages with less than 5 years of exposure to a GHS (compared to 12 percent among the control), and in 33 percent of those with 5 or more years.

For the quantity margin, a supply-side channel suggests several patterns. If the teachers' supply effect is important compared to parental demand, the impact of public high schools should be larger than that of public primary schools. The assumption is that while an adult with only primary level education is not a suitable teaching candidate (and indeed, 98 percent of teachers the LEAPS sample have at least secondary education), the difference in parental demand for primary education may not differ substantially across parents with primary versus secondary education. We thus consider the differential impact of public schools by level (primary versus high) and gender (boys versus girls - public schools are officially single sex). We expect high schools to be more important than primary schools, and, since teachers in private schools are primarily women, girl's high schools to have a larger impact than boy's high schools.

Instead of looking at schools, we can also focus directly on the supply of educated men and women. We look at whether there are more adult women with higher education in treatment villages (that received a girl's high school) and, in turn, whether the existence of a private school increases with fraction of adult women with higher education in the village (both absolutely and relative to adult men with higher education).

While the above patterns do lend further support to the supply-side channel, alternative explanations based on demand channels are possible: Several studies document the relative importance of women compared to men for investment in the human capital of their children, and secondary education could have a differential impact on parental demand compared to primary education. Therefore, we also test for the presence of the price effects mentioned above, that are harder to reconcile with a demand side explanation.

The price implications of the supply side and demand side channels differ. If the supply-side channel drives the existence of a private school, one would expect to see that the wage-bill for setting up a private school is lower in villages with pre-existing public girls' high schools. In the absence of such a school, the entrepreneur would either have to bring in teachers from outside (if possible) and pay for the compensating differential and travel costs; alternatively she would have to pay higher wages for educated male teachers. We can test whether our data support this hypothesis by looking at the wage-bill for of teachers in private schools.[8]

[8] An alternative would be to look at wages of secondary-school educated women who may or may not be teachers. The only available data source is the Pakistan Integrated Household Survey. Unfortunately, given

A complication is that we observe non-missing values of the wage-bill only in villages where private schools exist. To see why this poses a problem, consider the following hypothetical wage-bill distributions in villages with and without GHS.

$$[5, 6, 7, 8, 9, 10] \text{ (Without GHS)} \tag{7}$$
$$[3, 4, 5, 6, 7, 8] \text{ (With GHS)}$$

In the absence of any demand effects, suppose that private schools can only afford to set up in villages where the wage-bill is below 7. Thus, where we observe the wage-bill, $E(WB_i|WB_i$ is non missing, no GHS) - $E(WB_i|WB_i$ is non missing, GHS) $= 1$, which is lower than the uncensored $E(WB_i|$no GHS) - $E(WB_i|$GHS) $= 2$.

To deal with such potential selection issues we follow two approaches. We present results using a Heckman selection model, where the selection stage is the probability of observing a positive wage (this corresponds essentially to having a private school in the village - there are a few cases where a private school exists but wages are not reported). An alternative estimator is to use the "control-function" approach Angrist (1995), where we condition on the predicted probability of $WB_i = 1$ in the wage equation. Details of both approaches are in the Appendix.

In either approach, identification is based on the explicit non-linearity in the selection equation (see Duflo 2001 as an example). Augmenting the instrument set with potential candidates that are correlated to the probability of setting up a private school (where we observe wages) but uncorrelated to the wage-bill for entrepreneurs setting up a private school can help in identification and efficiency of the estimator. Following the industrial-organization work by Dowes and Greenstein (1996), we propose using the *number of public boys primary schools* as an additional instrument in the selection equation. In the presence of competitive effects, private schools should be less likely to setup in villages where there are boy's primary schools; additionally, such schools are unlikely to affect the wage-bill of the entrepreneur directly.

However, this instrument might be less than perfect. If boy's primary schools are differentially located in villages where the returns to education are low or if men and women compete

the small number of villages that received a GHS, the available sample sizes are too small—with the sample restrictions in our paper, we find only 3 villages in the treatment and 31 villages in the control set for these data.

directly for positions in the labor market for teachers, there could be a direct relationship between such schools and the wage-bill that entrepreneurs face. In our case, the additional instrument serves as a "robustness" check to the identification based only on non-linearities in the selection equation.

Finally, if there are simultaneous changes in the demand for schooling induced by girls' high schools, our estimate of the effect of GHS on the wage-bill will *always* represent a lower bound. To see this, reconsider our hypothetical example above and suppose the GHS increases the demand for private education, so that the minimum wage-bill at which private schools are profitable in villages with GHS is higher than those without GHS. In this case, $E(WB_i|\text{no GHS})$ - $E(WB_i|\text{GHS})$ recovers the joint-effect of a lower wage-bill and increases in demand—we cannot separate out the two effects without individual-level data.

# VI Results

## A. Baseline Differences

Recall our sample is the set of villages in patwar circles where no village had a GHS in 1981, the date of our baseline data. We define our treatment villages as those that received a GHS and control villages as those that either did not receive a girls' school at all, or those that received a girls' primary school. What did our treatment and control villages look like in 1981? Table II shows that in terms of female literacy, male literacy, gender ratios, and basic demographic attributes such as the percentage of male and female infants (less than age 4) and children (ages 5 to 14) there were no significant differences between treatment and control villages. Consistent with the stated rule, the population in treatment villages was almost twice as large as in the control. Using measures of infrastructure and public goods in these villages from the 1998 census shows little wealth differences (Appendix Table I). In fact judging from the percentage of households that own land in 1998, treatment villages fare slightly worse than control villages.

## B. OLS and First-Difference Specifications

Table III presents the results from our primary specification. The probit specification shows that a pre-existing girl's public secondary school increases the probability of a private school in the village by 9.7 percentage points (Column 1, Table 3). Since 12 percent of all control villages have a private school, this represents an 80 percent increase. An equally significant determinant of private school existence is village population; taken literally, the girl's secondary school effect is similar in magnitude to increasing village population by 2000 individuals (coincidentally, this is a one standard-deviation increase). The estimated impact remains significant at the 1 percent level when we introduce a full set of village level controls including exposure to other types of public schools, although the point-estimate is somewhat attenuated (Column 2, Table III). Introducing geographical dummies for patwar circles as in equation(3) increases the estimate and significance with magnitudes very similar to the first specification (Column 3, Table III).

Following equation(4), Columns (4) and (5) present results that control for time-invariant village effects by first-differencing the data at the village level. Interestingly, the effect of girl's high school on private school increases to almost 15 percentage points. This suggests some evidence of selection, although in the opposite direction to what one might expect: villages that received girl's high schools were also those where the returns to private schools were systematically lower. Thus not correcting for such selection would lead us to under-estimate the true effect of a girls high school. Comparing these villages to those that only received a girls' primary school yields identical results, suggesting that the selection effect is more than a pure competitive effect whereby pre-existing girls' schools decrease the residual demand for private education (results available with authors).

Column (5) presents further evidence of such negative selection by cluster specific time-trends in the first-difference specification. The GHS effect increases further to 17.4 percentage points. Since the specification is already in first differences, introducing cluster specific time-trends accounts for unobserved time-*varying* factors that are common across villages in the cluster and may affect both the likelihood of GHS provision and private school returns. Thus if a selection concern was that a village that received a road saw greater returns to education and hence greater demand for public (and private) schooling, as long as this road affected neighboring villages similarly, the estimate of GHS on private schools is unbiased and consistent.

Non-parametric (propensity-score) techniques yield similar results. A pre-existing girl's high school increases private school existence probabilities by 11 to 14 percentage points depending on whether we use local linear regression or kernel matching (results available with authors).

These results suggest two patterns: if time-varying omitted-variables play a small role, GHS are critical for the future development of private schools, both absolutely and relative to other types of schools. Furthermore, the difference in the estimated coefficient using the cross-section and panel lead us to believe that GHS were not distributed randomly—they were systematically targeted to villages where private schools were less likely to emerge. The next section explores whether selection on time-varying characteristics was also important.

## C. Instrumental Variable Specifications

### C-1. First-Stage Results

Table IV, Columns (1-3) present a series of first stage regressions using our intention-to-treat variable as a predictor for the location of GHS. Without additional controls, a village with population above 500 that was ranked first in population terms within the patwar circle (and thus in our intended-to-treat group), was 6.4 percentage points more likely to receive a girl's high school (Column 1, Table IV). Since 4.7 percent of all villages received such a school, this is close to an increase of 130 percent.

Part of this is a population effect whereby larger villages are likely to rank high within the patwar circle and likely to receive a girl's high school. In Column 2, we condition on linear and quadratic terms of the village's population in 1981 and the maximum village population in the patwar circle in 1981. Although the point-estimate is reduced by half, it remains large and highly significant: villages that satisfied the construction rule under the SAP were 55 percent more likely to receive a girl's high school than those that did not. We obtain very similar results with a more exacting first stage that includes a full set of location dummies for administrative units known as "qanoongho halqa" (QH), which include around 10 PCs (Column 3, Table 4).[9]

Our instrument performs fairly well in the first-stage regression. Following Stock, Wright and Yogo (2002), a useful characterization of the strength of an instrument is the "concentration parameter", inferences for which can be based on the F-statistic from the first-stage. For all three specifications—without additional controls, with population controls and with additional geographical dummies—the F-statistic is greater than 14 and exceeds the proposed critical

[9] We cannot use PC fixed effects since our instrument and controls rely on PC level population measures (i.e. the maximum village population in a PC).

thresholds (approximately 9) for testing the null hypothesis that the instruments are weak (Stock, Wright and Yogo 2002). Moreover, the explicit conditioning on polynomial population terms implies that the remaining variation induced by the instrument is non-monotonic and non-linear and therefore likely un-correlated with omitted variables in our primary specification.

### C-2. Instrumental Variables Estimates

Columns (4) to (5) present the corresponding linear IV coefficients. The coefficient increases dramatically and the significance drops to the 10 percent level. While part of this increase can be attributed to selection on time-varying omitted variables, we think it unlikely that these effects can be as large as the estimates suggests. Column (6) assesses whether functional specification plays a role. We implement a bivariate probit specification as in equation(5) and report the marginal impact of girl's high school on the existence of a private school. The standard errors are bootstrapped at sample values of other variables (alternative standard errors calculated at the mean of the sample-value for other variables yield similar results). The point-estimate from the bivariate probit are half those of the linear IV and are significant at the 5 percent level. The estimate suggests (negative) selection on unobservable variables, and is double what we obtain with the first-differenced specification: taken at face-value, constructing a girl's secondary school increases the probability of a private school in the village by 36 percentage points, or over 300 percent.

Why are there large differences between the biprobit and the linear IV estimator? While the instrument has high predictive power in the first-stage, the probability of the treatment is low. A heuristic argument for the large difference is that the linear IV projects the change in treatment probabilities resulting from the instrument to the entire sample, while the bivariate probit, by imposing a normal structure on the error, accounts for the differential change in probabilities from the same change in the index value at different points in the distribution. This is particularly a problem with low treatment values, where the cumulative distributions of the uniform and normal differ substantially.

One way to understand how large these problems are likely to be is through Monte-Carlo simulations under similar conditions. In independent simulations, we consider various treatment (T) and outcome (Y) probabilities in a Monte-Carlo setting and show that the difference between linear IV and biprobit estimates can be understood in the context of a 2 X 2 matrix where T={0,1} depict the rows and Y={0,1} the columns. When individual cell-proportions are very different from a 1/4th division of the entire sample, linear IV and biprobit estimates differ, and the problem becomes worse when the deviations are larger. In the context of our problem, these deviations are indeed large with 87 percent of the sample concentrated in the single cell, T=0 and Y=0.

Encouragingly, we also find that in the presence of negative selection (that is, the naive OLS estimator is biased towards zero), the biprobit estimator is never below the OLS estimate: as long as IV> biprobit>OLS, the OLS estimate provides a lower-bound for the true treatment effect. Thus, while we cannot precisely pin down a point-estimate for the effect of GHS on private school existence, the IV and biprobit results confirm that there is a strong, positive causal impact that is bounded below by the first-difference estimate of 18 percentage points.

The differences in the OLS and IV results also implies that the villages chosen to receive a GHS were those where private schools were systematically less likely to locate. Since baseline characteristics were identical in our treatment and control villages, the selection was based on unobserved time-varying village effects. One interpretation—advanced for instance, by Pitt, Rosenzweig and Gibbons (1995) in their study of program placement in Indonesia—is that the government acts altruistically, trying to equalize differences between villages. Villages where the demand for education was less likely to increase received GHS; these were precisely the villages where private schools were less likely to locate.

We are somewhat sceptical of this altruistic argument, particularly given the nature of the state in Pakistan's history. In talking to education department officials who had been in the ministry at the time of the program, it was evident that apart from the SAPP rules (which we were shown in a number of offices), local-level politics also played a role. A more critical explanation is that these schools were targeted to villages with powerful local landlords and officials. The context in Pakistan suggests that these are precisely the villages where the demand for education is lower, and less likely to increase over time. Construction in villages with a lower demand for education could reflect political-economy considerations rather than a desire for equity. Finally, the selection could also have been on wages—if villages that received a girl's high school were also those that experienced high wage growth, this would also deter the entry of private schools. This seems unlikely. As we show below, wages in villages with a GHS are *lower* than in others.

## D. Potential Channels

Tables V-VI presents results from a set of estimations based on our previous discussion. We examine four questions: (a) what is the relative importance of girl's high schools for private school existence as compared to other types of public schools?; (b) what was the contribution of girl's high schools to the supply of educated women; (c) what is the contribution of educated females to the existence of a private school, and (d) is there a relationship between wages and the construction of a girl's high schools. These results lend support to a supply-side channel.

If private schools arise because of the availability of "women as teachers", we expect a larger impact of GHS compared to other types of public schooling. Columns (1)-(2) in Table V present estimates from a probit and linear probability model, where the latter includes geographical dummies for the patwar-circle that the village is in. Both specifications confirm the importance of GHS relative to other types of public schooling, with coefficients for years of exposure to a girls' high school almost three times as large an effect as that of the next most important public school type.

Since selection effects are important, Columns (3) and (4) present results from a first-difference specification with and without cluster-specific time-trends. These results magnify the importance of GHS: the change (from 1981 to 1998) in whether a village has a GHS or not is the *only* schooling variable that matters, and the magnitude of the effect is large. Whether a village received a boys' primary/high or girls' primary school between 1981 and 1998 has no affect on the likelihood of a private school setting up in the village. These results suggest the supply side channel is more likely since they constrain the routes through which a demand-side story can work: it must be the case that the father's education does not stimulate demand for their children's education and also that primary schooling for mothers is not enough. The only mothers who demand more education for their children are those who have eight or more years of schooling—indeed, there must be strong nonlinearity in the demand for children's education and mother's schooling.

Columns (5)-(8) present the next logical step. We assess the correlation between educated women and the presence of a GHS for a variety of specifications. In all specifications, a GHS increases the percentage of adult women with higher levels of education (equal to or more than 8 years of schooling) by 1.5-2.2 percentage points and the estimated increases are significant at the 1 percent level of confidence. While this may appear as a small effect, only 1.3 percent of all women in the average village in 1981 had higher levels of education. Therefore receiving a girl's high school more than doubles this percentage.

Columns (9)-(12) in Panel B examines the importance of highly educated women for the existence of a private school. Are such women more important than similarly educated men? Restricted mobility and employment opportunities for women suggests that educated women should have a bigger impact than educated men, since their wage rates will be lower. This is indeed the case. While the effect of educated men is only slightly smaller in the basic probit specification in column (9), this difference between educated men and women increases substantially once we control for geographical location, suggesting that part of the estimated coefficient on male education was due to omitted geographical characteristics. For our preferred first-difference specification, the impact of women with 8 or more years of schooling remains as strong while the percentage of educated males has *no* impact on the existence of a private school.

Table VI then examines the wage bill across villages. Ideally we would like to have data on the wages of women with high school education (i.e. the level necessary to become primary school teachers) in all villages. However, there are few sources that provide such village level wage estimates. Moreover, in a lot of cases there are few or nonexistent paid labor markets for women, especially those with high school education, in these villages. As a result we will use the private school census (PEIP) that includes teacher wages in all private schools in the village. While these wages cannot be separated by gender, since the vast majority of private school teachers are women, the wage bill is likely to reflect wages to skilled women.

Column (1) presents the simple OLS results in the restricted sample where we have wage data.[10] The results are large and significant and in the direction predicted by the supply side

[10] This is slightly smaller than the number of villages where there is a private school since in a few cases in the PEIP data private schools did not report wages.

channel. Private schools in villages which have a GHS have an 18% lower wage-bill. Moreover, to the extent that we face attenuation bias arising from a noisy measure of women's wages (average wages in private schools), the actual cost differential may be even higher.

Columns (2)-(5) correct for the fact that we do not observe wages for our full sample. Columns (2)-(3) present results from estimating a Heckman selection model and Columns (4)-(5) present the alternate "control function" approach (see Appendix). Columns (2) and (4) identify solely of non-linearities, whereas Columns (3) and (5) introduce an additional instrument for the selection stage - the number of government boys' primary schools in the village. The results are very similar to the OLS estimates suggesting that selection into the non-zero wage sample is not a serious issue. Interestingly as a useful falsification exercise, note that pre-existing boy high schools have an insignificant though positive effect, suggesting that high school exposure is not picking up some other effect and, to the extent that the average wage bill is also influenced by male teacher wages, that labor markets may indeed not be as localized for men.

Together with the results on the quantity margin - that GHS and not other forms of public schooling, and female higher education and not male higher education, affect private school existence, the wage-bill results present direct evidence that the supply-side "women-as-teachers" channel is indeed important in determining where private schools are setup. Given localized labor markets for women, GHS reduce wages in the village for educated women by increasing the relative supply of such women, and this in turn allows private school to exist by taking advantage of an affordable local supply of teachers.

# VII Conclusion

With 2 million adolescents poised to enter the labor-force each year, human capital investments will play a large role in determining where these children will fit into a fast globalizing world. Our parallel work suggests that children in rural private schools will have learnt a lot more by the time they drop-out of the formal schooling system than their counterparts (Andrabi and others 2005b). At the same time, distance plays a role—children will attend private schools only if they are physically located in the village where they reside. Fostering greater private school construction in rural areas is critical.

The causal impact of girl's high schools on subsequent location of a private school calls for public investment-led growth in human capital, in the spirit of "big-push" arguments advanced by Rodenstein-Rodan (1943) and Murphy, Shleifer and Vishny (1989). In contrast to the literature that calls for larger primary school compared to secondary school investment, our findings suggest that both play a role. That the students in today's schools are the potential repositories of human capital for the next generation implies that low-income countries can enter a "virtuous cycle" by investing heavily in the creation of a cohort of educated women.

The results on *how* girl's high schools lead to the creation of private schools in the next generation are a testimony to the resilience of the private sector and its ability to convert cultural constraints—the restricted mobility and labor-market opportunities for women—into an advantage. Villages with girl's high schools are also those with a larger stock of educated women, who can then teach in private schools. With limited mobility, a private school entrepreneur becomes a virtual monopsonist when located in such a village. At one level, this seems like a fairly pernicious outcome: women receive far lower wages in the labor market compared to men for the same job. At the same time, these cost savings are directly passed on to the children who study in these schools; in the absence of this labor market distortion, it is unclear whether these schools would have arisen in rural areas, and if so, whether they could have catered to the children from middle-income families that are currently enrolled in them.

The results also provide a fascinating glimpse of education in high-income countries during the early to mid-twentieth century. Recent papers concerned about the decline in high-quality teachers in US public schools point to greater employment opportunities and increasing pay for women in non-teaching fields (although Hoxby 2004 finds that wage-compression due to unionization played a larger role than increasing competition from other employers). The rise of private schools in Pakistan suggest that in low-income countries at least, the "implicit-subsidy" to education from low female labor-force participation is alive and kicking.

# Appendix

**Dealing with Selection Issues in the Wage Bill**

Since we only observe the wage bill in villages where there is a private school, a concern described in the main text is that simple OLS estimates may be biased if such selection is not accounted for. Here we provide details on two approaches we use in the paper to address such concerns.

Following Angrist (1995), the problem can be formally stated as follows.

The wage-bill is determined through a linear equation conditional on the existence of a private school

$$WB_i = \alpha + \beta GHS_i + \varepsilon_i \tag{8}$$

and a censoring equation (denoting $WB_i = I$ as the indicator for whether $WB_i$ is non-missing)

$$WB_i = I\{\delta GHS_i - \nu_i > 0\} \tag{9}$$

The instrument $Z_i$ determines a first stage

$$GHS_i = \gamma + \mu Z_i + \tau_i \tag{10}$$

Given the validity of the instrument, $Z_i$, we assume that $cov(\tau_i, Z_i) = 0$. Then,

$$E(\varepsilon_i|Z_i, WB_i = 1) = E(\varepsilon_i|Z_i, (\delta\gamma + \delta\mu Z_i > \nu_i - \delta\tau_i)$$

so that $cov(\varepsilon_i, Z_i) \neq 0$ in equation(8) above. Thus, although $Z_i$ is a valid instrument for the decision to setup a private school, it is not a valid instrument in equation(8). There are two potential solutions.

Following Heckman (1978) if we are willing to assume that $(\varepsilon_i, \nu_i, \tau_i)$ are jointly normally distributed, homoskedastic and independent of $Z_i$, we obtain the familiar "mills-ratio" as the relevant expectation function conditional on participation. That is,

$$\begin{aligned} E(\varepsilon_i|Z_i, (\delta\gamma + \delta\mu Z_i &> \nu_i - \delta\tau_i) = \\ E(\varepsilon_i|Z_i, (\delta\gamma + \delta\mu Z_i &> \nu_i - \delta\tau_i) = \lambda(\delta\gamma + \delta\mu Z_i) \end{aligned}$$

where $\lambda(\delta\gamma + \delta\mu Z_i) = \frac{-\phi(\lambda(\delta\gamma+\delta\mu Z_i))}{\Phi(\lambda(\delta\gamma+\delta\mu Z_i))}$ and $\phi(.)$ and $\Phi(.)$ are the density and distribution functions of the normal distribution for $\nu_i - \delta\tau_i$. This mills-ratio can is then directly included in equation(8) as the appropriate selection-correction.

An alternative approach, proposed by Heckman and Robb (1986) and developed by Ahn and Powell (1993) uses the "control-function" approach, where we condition on the predicted probability of $WB_i = 1$ in equation(8). In essence, this method proposes to estimate $\beta$ by using pair-wise differences in $WB_i$ for two villages (in our case) for which the non-parametric probability of participation is very close. The approach is implemented by first estimating equation(9) directly, and then including the predicted probability of participation (and its polynomials) as additional controls in equation(8).

## References

[1] Abadie, A. 2003. "Semiparametric instrumental variable estimation of treatment response models," *Journal of Economterics*, Vol. 113, 231-263.

[2] Ahn, H. and J.L. Powell. 1993. "Semi-parametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," Journal of Econometrics 58, 3-29.

[3] Altonji, Joseph, Todd Elder and Christoper Taber 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools" forthcoming *Journal of Political Economy.*

[4] Andrabi, Das and Khwaja 2005a. "Private Schooling in Pakistan: Possibilities and Limits." Working Paper.

[5] Andrabi and others 2005b. "Report on Learning and Educational Attainments in Pakistani Schools. Working Paper.

[6] Andrabi, T., J. Das, A. Khwaja and T. Zajonc, 2005c, "Religious Education in Pakistan: A Look at the Data", John F. Kennedy School of Government, Faculty Research Working Paper Series, Working Paper Number RWP05-024.

[7] Angrist, J. and V. Lavy, 1999. "Using Maimonedes Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, Vol. 114, No. 2, 533-575.

[8] Angrist, J. 1995."Conditioning on the Probability of Selection to Control Selection Bias," NBER Technical Working Paper No. 181.

[9] Alderman, Harold J.Kim and Peter Orazem.1990. "Design, Evaluation and Sustainability of Private Schools for the Poor: The Pakistan Urban and Rural Fellowship School Experiments." *Economics of Education Review* 22 (2003): 265-274

[10] Besley, Timothy, and Stephen Coate, "Public Provision of Private Goods and the Redistribution of Income." The *American Economic Review* 81(4), 979-984. September 1991.

[11] Campbell, D.1969. "Reforms as Experiments," *American Psychologist*, XXIV, 409-429.

[12] Downes, Thomas A. and Shane M. Greenstein. 1996. "Understanding the Supply Decisions of Nonprofits: Modeling the Location of Private Schools" *Rand Journal of Economics* Vol. 27(2)

[13] Downes, Thomas A. & Greenstein, Shane M. 2002. "Entry into the Schooling Market: How is the Behaviour of Private Suppliers Influenced by Public Sector Decisions?" *Bulletin of Economic Research* 54 (4), 341-371.

[14] Dulfo, E. (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," American Economic Review, Vol. 91, No. 4.

[15] Government of Punjab, Department of Education. 2002. "Educational Management Information System for Punjab."

[16] Government of Pakistan, 1983, Sixth Five Year Plan 1983-88, Planning Commission, Islamabad

[17] Heckman, J. J., and R. Robb. 1986. "Alternate Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in H. Wainer, ed., Drawing Inferences from Self-Selected Samples, New York: Springer-Verlag.

[18] Hoxby, Caroline M. 1994. "Do Private Schools Provide Competition for Public Schools?" National Bureau of Economic Research Working Paper Number 4978.

[19] Hoxby, Caroline M. and Andrew Leigh. 2004. "Pulled Away or Pushed Out? Explaining the Delhine of Teacher Aptitude in the United States." *American Economic Review*, Vol.93(2)

[20] Jimenez, Emmanuel and Jee-Peng Tan. 1985. "Educational development in Pakistan : the role of user charges and private education, Volume 1." Education and training series Discussion Paper Number EDT 16. World Bank. Washington, D.C.

[21] Jimenez, Emmanuel and Jee-Peng Tan. 1987. "Decentralized and private education : the case of Pakistan, Volume 1." Education and training series Discussion Paper Number EDT 67. World Bank. Washington, D.C.

[22] Jimenez, Emmanuel, Marlaine E Lockheed and Vicente Paqueo. 1999. "The relative efficiency of private and public schools in developing countries." *World Bank Research Observer.* Vol. 6, no.2 (July 1991), pp. 205-218.

[23] Kingdon, Geeta. 1996. "The Quality and Efficiency of Private and Public Education: A Case-study of Urban India." *Oxford Bulletin of Economics and Statistics.*

[24] Stock, James H., Jonathon H. Wright and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." Journal of Business and Economic Statistics, Vol.20(4) 518:529.

[25] Tooley, James. "Private Schools for the Poor." *Education Next.* Fall 2005: 22-33.

[26] Murphy, Kevin M & Shleifer, Andrei & Vishny, Robert W, 1989. "Industrialization and the Big Push," *Journal of Political Economy*, University of Chicago Press, vol. 97(5), pages 1003-26.

[27] Government of Pakistan, 1997, Report on the PIHS 1995-96, Islamabad

[28] Pitt, Mark M., Mark R. Rosenzweig and Donna M. Gibbons, 1995. "The Determinants and Consequences of the Placement of Government Programs in Indonesia." In D. van de Walle and K. Nead, eds., *Public Spending and the Poor: Theory and Evidence*. Baltimore: Johns Hopkins University Press for the World Bank.

[29] Psacharopolous, G. 2002. "Returns to Investment in Education: A Further Update." World Bank Policy Research working paper series Number 2881. World Bank, Washington D.C.

[30] Rodenstein-Rodan, Paul. 1943. "Problems of Industrialization of Eastern and South-Eastern Europe." *Economic Journal* 53 (210-211): 202-211.

[31] Tilak, Jandhyala B.G. and Ratna M. Sudarshan. 2001. "Private Schooling in Rural India." National Council of Applied Economic Research Working Paper #WP010009. NCAER, New Delhi, India.

[32] World Bank. 2005. "Improving Gender Outcomes: The Promise for Pakistan.". World Bank, Washington, D.C.

**TABLE I**
SUMMARY STATISTICS

| variable | mean | sd | N |
|---|---|---|---|
| **Years Exposure - GHS** | 0.68 | 3.23 | 6968 |
| **Years Exposure - GPS** | 7.19 | 7.24 | 6960 |
| **Years Exposure - BHS** | 6.34 | 20.35 | 6960 |
| **Years Exposure - BPS** | 26.73 | 20.39 | 6869 |
| **Private School Exists?** | 0.13 | 0.34 | 6968 |
| **Number of Private Schools** | 0.22 | 0.81 | 6968 |
| **1998 % Enrolled in Private Schools** | 0.10 | 0.21 | 902 |
| **1998 Population** | 1829.09 | 2023.31 | 6968 |
| **1981 Population** | 1210.50 | 1272.31 | 6968 |
| **1998 Number of Women w/ Matric and Above Education** | 13.07 | 39.36 | 6968 |
| **1998 % HHs w/ Permanent Housing** | 0.07 | 0.05 | 6723 |
| **Village Land Area** | 1647.79 | 2340.71 | 6874 |
| **Number of Villages in Patwar Circle** | 4.38 | 2.12 | 6968 |

Summary statistics are for the sample of villages that (a) did not have girls' high or primary school prior to 1981 and (b) villages whose neighbors did not have a girls' high school before 1981. Land is measured in Kanals.

**TABLE II**

BASELINE DIFFERENCES IN MEANS

| | Treated | Not Treated | Difference |
|---|---|---|---|
| **Number of Villages** | 328 | 6640 | |
| **1981 Female Literacy Rate** | 0.017 | 0.015 | 0.002 |
| | (0.007) | (0.001) | (0.007) |
| **1981 - % adult women with Middle and above Education** | 0.016 | 0.012 | 0.004 |
| | (0.007) | (0.001) | (0.007) |
| **1981 % girls age 0-4** | 0.154 | 0.155 | -0.001 |
| | (0.020) | (0.004) | (0.020) |
| **1981 % girls age 5-14** | 0.289 | 0.285 | 0.004 |
| | (0.025) | (0.006) | (0.026) |
| **1981 adult Male Literacy Rate** | 0.184 | 0.164 | 0.020 |
| | (0.021) | (0.005) | (0.022) |
| **1981 - % adult men with Middle and above Education** | 0.135 | 0.116 | 0.019 |
| | (0.019) | (0.004) | (0.019) |
| **1981 % boys age 0-4** | 0.143 | 0.143 | 0.001 |
| | (0.019) | (0.004) | (0.020) |
| **1981 % boys age 5-14** | 0.295 | 0.293 | 0.002 |
| | (0.025) | (0.006) | (0.026) |
| **1981 Female/Male Ratio** | 0.911 | 0.906 | 0.005 |
| | (0.016) | (0.004) | (0.016) |
| **1981 Population** | 2069.69 | 1168.05 | 901.6324*** |
| | (94.17) | (15.12) | (71.16) |

The table shows baseline differences between treatment and control villages. Standard-errors of t-tests or proportion tests (as appropriate) are in parenthesis.

**Table III - Private School Existence and Previous Girls High Schools**

| | (1) Probit | (2) Probit - All controls | (3) OLS (PC Location Dummies) | (4) First diference | (5) First diference & PC Dummies |
|---|---|---|---|---|---|
| Treatment- Received GHS | 0.097 | 0.0646 | 0.0928 | 0.1494 | 0.1739 |
| | (0.0223) | (0.0207) | (0.0247) | (0.0250) | (0.0241) |
| 1998 Population (000s) | 0.051 | 0.0391 | 0.0905 | | |
| | (0.0032) | (0.0075) | (0.0176) | | |
| 1998 Population (000s) Sq | -0.0014 | -0.0011 | -0.0046 | | |
| | (0.0002) | (0.0003) | (0.0014) | | |
| 1981 Population (000s) | | 0.0275 | 0.0134 | | |
| | | (0.0133) | (0.0281) | | |
| 1981 Population (000s) Sq | | -0.0013 | 0.0029 | | |
| | | (0.0012) | (0.0041) | | |
| % Perm Houses | | 1.2862 | 0.9383 | | |
| | | (0.0821) | (0.1804) | | |
| 1998-1981 Population (000s) | | | | 0.0795 | 0.1162 |
| | | | | (0.0070) | (0.0079) |
| Years Exposure - GHS | | | | | |
| Years Exposure - GPS | | 0.001 | -0.0001 | | |
| | | (0.0005) | (0.0007) | | |
| Years Exposure - BPS | | 0.0001 | 0.0004 | | |
| | | (0.0002) | (0.0003) | | |
| Years Exposure - BHS | | 0.0011 | 0.002 | | |
| | | (0.0002) | (0.0003) | | |
| With Patwar-Circle Dummies | NO | NO | YES | | |
| With PC cluster-specific time trends | | | | NO | YES |
| Observations | 6968 | 6761 | 6761 | 6968 | 6968 |
| Pseudo R-sq | 0.1 | 0.18 | | | |
| Adj R-sq | | | 0.34 | 0.07 | 0.3 |

The table shows the relationship between the existence of a private school and GHS. Columns (1) and (2) estimate non-linear probability models (probit) and column (3) the corresponding linear specification. Columns (4) and (5) present results from the village-level first-differenced specification. Robust standard errors in parentheses.

**Table IV - Private School Existence - Instrumental Variables**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | First- Stage Probit | First-Stage Probit | First-Stage (QH Location Dummies) | Linear 2nd-Stage | Linear 2nd-Stage- QH Location Dummies | BiProbit - Boostrapped SEs (xx vars are also included but Cioeffs and Ses not reported) |
| Girls High School Rule | 0.0642 | 0.026 | 0.0297 | | | |
| | (0.0067) | (0.0079) | (0.0078) | | | |
| Treatment- Received GHS | | | | 0.8277 | 0.7507 | 0.3567 |
| | | | | (0.4808) | (0.4368) | (0.1331) |
| 1981 Population (000s) | | 0.0268 | 0.0335 | 0.0269 | 0.0194 | xx |
| | | (0.0059) | (0.0066) | (0.0263) | (0.0252) | |
| 1981 Population (000s) Sq | | -0.0023 | -0.0017 | -0.0016 | -0.0008 | xx |
| | | (0.0006) | (0.0008) | (0.0015) | (0.0015) | |
| 1981 Max Population (000s) in PC | | -0.0025 | 0.0065 | -0.0041 | -0.0061 | xx |
| | | (0.0049) | (0.0063) | (0.0089) | (0.0094) | |
| 1981 Max Population (000s) sq in PC | | 0.0005 | 0.0002 | 0.00001 | -0.0004 | xx |
| | | (0.0005) | (0.0008) | (0.0014) | (0.0012) | |
| 1998 Population (000s) | | | | 0.0403 | 0.0544 | xx |
| | | | | (0.0105) | (0.0099) | |
| 1998 Population (000s) Sq | | | | -0.0001 | -0.0009 | xx |
| | | | | (0.0005) | (0.0005) | |
| % Perm Houses | | | | 1.3055 | 0.7645 | xx |
| | | | | (0.1053) | (0.1516) | |
| Observations | 6968 | 6968 | 6968 | 6874 | 6874 | 6874 |
| Chi-sq/F-Test (GHS Rule = 0) | 122.61 | 12.94 | 14.67 | | | |
| Pseudo R-sq | 0.05 | 0.07 | | | | |
| Number of QGH 1998 | | | | | 656 | |
| Prob > chi2 | 0 | 0 | | | 0 | 0 |
| Prob > F | | | 0 | 0 | | |
| Adj R-sq | | | 0.07 | | | |

The first three columns in the table show the first-stage of the IV strategy. Column (1) shows the bivariate correlation between the eligibility rule an GHS. Columns (2) and (3) are the corresponding first-stages for Columns (4) and (5); Column (6) reports the estimated marginal impact of GHS and bootstrapped standard-errors for a bivariate probit specification (xx represents variables included in the regression, but whose marginal coefficients and standard errors we have not bootstrapped for computational convenience). Standard errors in parentheses.

**Table V - Private School Existence - The Female Teacher Channel?**

**PANEL A**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | Dependent Variable: Private School Existence | | | | Dependent Variable: Percentage of Women with Middle and Above Education | | | |
| | Probit | OLS - Controls & PC Location Dummies | First Difference | First diference & PC Dummies | OLS | OLS- Controls & PC Location Dummies | First Difference | First diference & PC Dummies |
| Years Exposure - GHS | 0.0044 | 0.0059 | | | | | | |
| | (0.0010) | (0.0016) | | | | | | |
| Years Exposure - GPS | 0.0016 | -0.0002 | | | | | | |
| | (0.0006) | (0.0007) | | | | | | |
| Years Exposure - BHS | 0.0013 | 0.002 | | | | | | |
| | (0.0002) | (0.0003) | | | | | | |
| Years Exposure - BPS | 0.0002 | 0.0004 | | | | | | |
| | (0.0002) | (0.0003) | | | | | | |
| Treatment- Received GHS | | | | | 0.0221 | 0.015 | 0.015 | 0.0183 |
| | | | | | (0.0037) | (0.0042) | (0.0031) | (0.0039) |
| 1998-1981 Population (000s) | | | 0.0798 | 0.116 | | | -0.0014 | 0.0039 |
| | | | (0.0071) | (0.0081) | | | (0.0012) | (0.0013) |
| Change in Exposure - GHS | | | 0.1515 | 0.16 | | | | |
| | | | (0.0255) | (0.0250) | | | | |
| Change in Exposure - GPS | | | 0.0103 | -0.008 | | | | |
| | | | (0.0081) | (0.0107) | | | | |
| Change in Exposure - BHS | | | -0.0645 | -0.0314 | | | | |
| | | | (0.0438) | (0.0693) | | | | |
| Change in Exposure - BPS | | | -0.0144 | -0.0126 | | | | |
| | | | (0.0088) | (0.0114) | | | | |
| Location Dummies | NO | YES | NO | | NO | YES | NO | |
| Cluster-Specific Time-Trends | NO | NO | NO | YES | NO | NO | NO | YES |
| Observations | 6854 | 6761 | 6854 | 6854 | 6967 | 6767 | 6964 | 6964 |
| Pseudo R-sq | 0.12 | | | | | | | |
| Adj R-sq | | 0.34 | 0.07 | 0.3 | 0.01 | 0.5 | 0.003 | 0.38 |

**PANEL B**

| | (9) | (10) | (11) | (12) |
|---|---|---|---|---|
| | Dependent Variable: Private School Existence | | | |
| | Probit | Controls & PC FEs | First Difference | First diference & PC Dummies |
| % middle & above females | 0.4149 | 0.52 | | |
| | (0.0819) | (0.1217) | | |
| % middle & above males | 0.3506 | 0.0783 | | |
| | (0.0469) | (0.0738) | | |
| Change in % Females middle+ | | | 1.0146 | 0.5801 |
| | | | (0.1029) | (0.1153) |
| Change in % Males middle+ | | | 0.0498 | -0.0118 |
| | | | (0.0531) | (0.0716) |
| 1998-1981 Population (000s) | | | 0.0839 | 0.1186 |
| | | | (0.0076) | (0.0080) |
| Observations | 6967 | 6873 | 6964 | 6964 |
| Pseudo R-sq | 0.17 | | | |
| Prob > chi2 | | | | |
| Adj R-sq | | 0.34 | 0.09 | 0.3 |

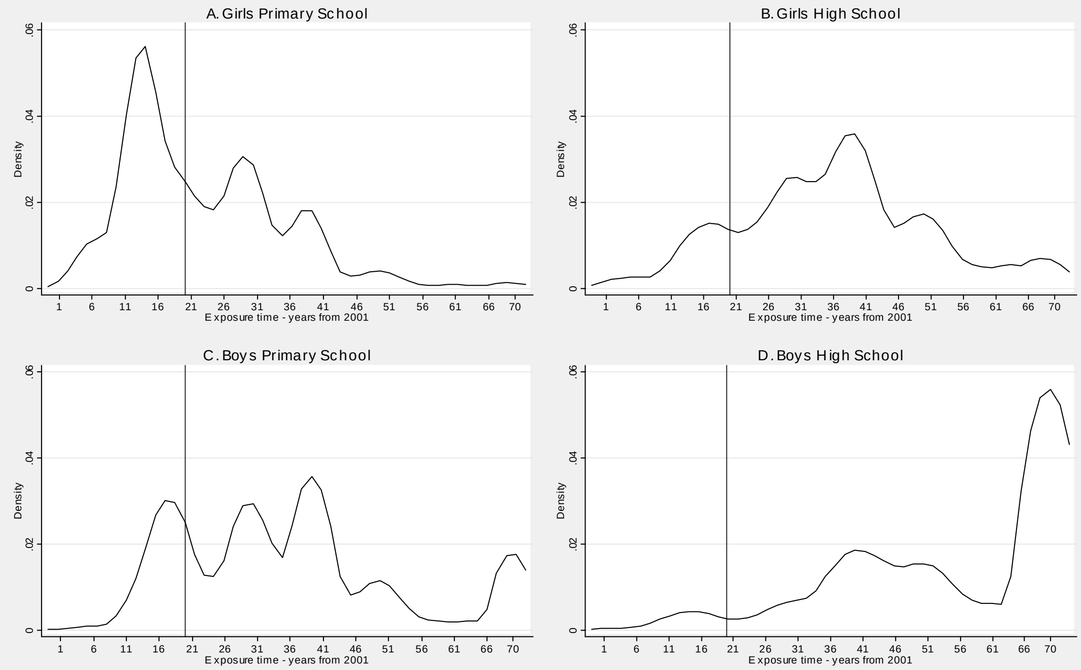
Columns (1) to (4) examines the relationship between different types of government schools and private school existence. Column (1) is a probit, Column (2) a linear specification with location dummies; Columns (3) and (4) are the village-level first-difference. Columns (5) to (8) look at the impact of GHS on female higher education. Columns (9) to (12) looks at private school existence and female/male education.

**Table VI - Supply Side Impact - Teaching Costs**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | OLS - Controls & QH Dummies | Heckman- Controls & QH Dummies | Heckman - Controls & QH Dummies, BPS | Control Function Controls & QH Dummies | Control Function Controls & QH Dummies, BPS |
| Treatment- Received GHS | -0.1977 | -0.2015 | -0.2037 | -0.2031 | -0.2094 |
| | (0.1078) | (0.0790) | (0.0794) | (0.1079) | (0.1083) |
| Years Exposure - BHS | 0.0006 | 0.0004 | 0.0004 | 0.0002 | 0.0002 |
| | (0.0010) | (0.0008) | (0.0008) | (0.0011) | (0.0011) |
| 1998 Population (000s) | 0.0329 | 0.0022 | 0.0127 | -0.0161 | -0.0047 |
| | (0.0233) | (0.0320) | (0.0309) | (0.0451) | (0.0433) |
| 1998 Population (000s) Sq | -0.0004 | 0.0004 | 0.0001 | 0.0010 | 0.0007 |
| | (0.0010) | (0.0011) | (0.0010) | (0.0015) | (0.0014) |
| Observations | 877 | 6967 | 6967 | 877 | 877 |
| Pseudo R-sq | | | | | |
| Prob > chi2 | | 0 | 0 | | |
| Adj R-sq | 0.15 | | | 0.15 | 0.15 |

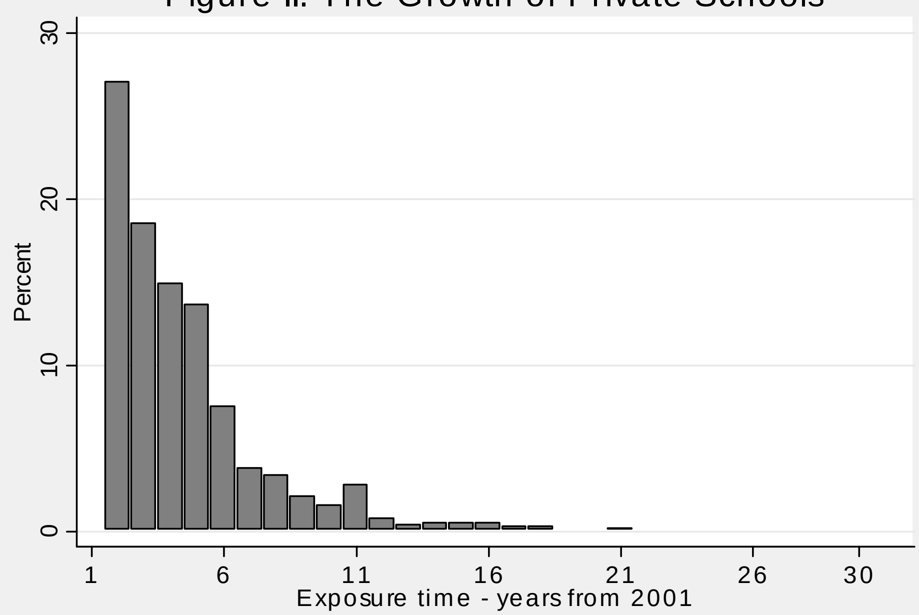
Columns (1) to (5) examines the relationship between average wage bill in private schools and government high schools. Column (1) runs an OLS specification. Columns (2)-(3) run a Heckman selection model to take into account the fact that the LHS variable is only observed in villages where private schools exists. Column (3) differs in that it includes an additional instrument for the selection stage - the number of government boys primary schools. Columns (4)-(5) present an alternate "control function" method to account for the selection issue by directly including polynomials in the predicted probability of observing a positive wage in the wage regression. Column (5) differs in that it includes an additional instrument for the selection stage - the number of government boys primary schools.

# Figure I. Government School Exposure
conditional on receiving school




Schools more than 70 years old are coded as 70

# Figure II. The Growth of Private Schools



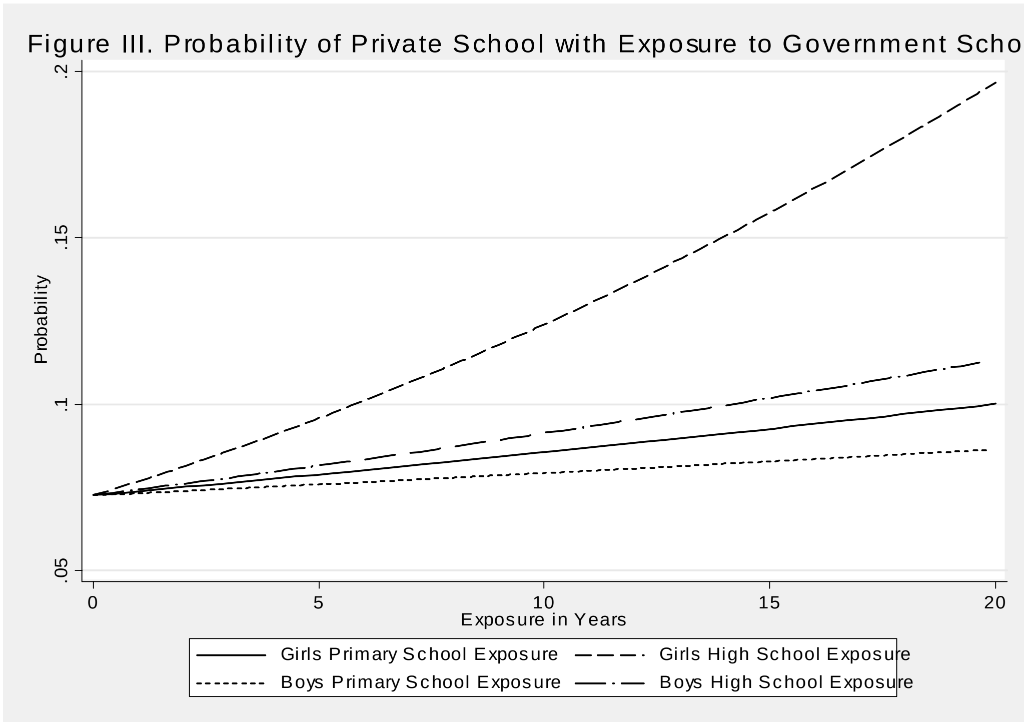


Figure III. Probability of Private School with Exposure to Government Schoo
Probability
.05
.1
.15
.2
0
5
10
15
20
Exposure in Years
Girls Primary School Exposure
Girls High School Exposure
Boys Primary School Exposure
Boys High School Exposure

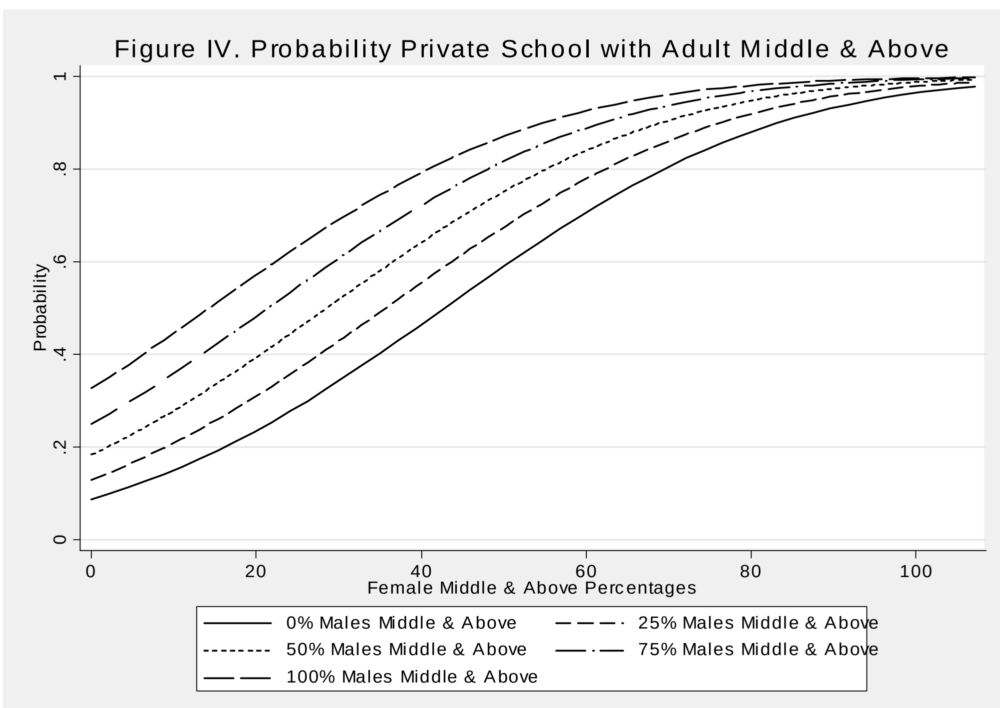



Figure IV. Probability Private School with Adult Middle & Above
Probability
0
.2
.4
.6
.8
1
0
20
40
60
80
100
Female Middle & Above Percentages
0% Males Middle & Above
25% Males Middle & Above
50% Males Middle & Above
75% Males Middle & Above
100% Males Middle & Above

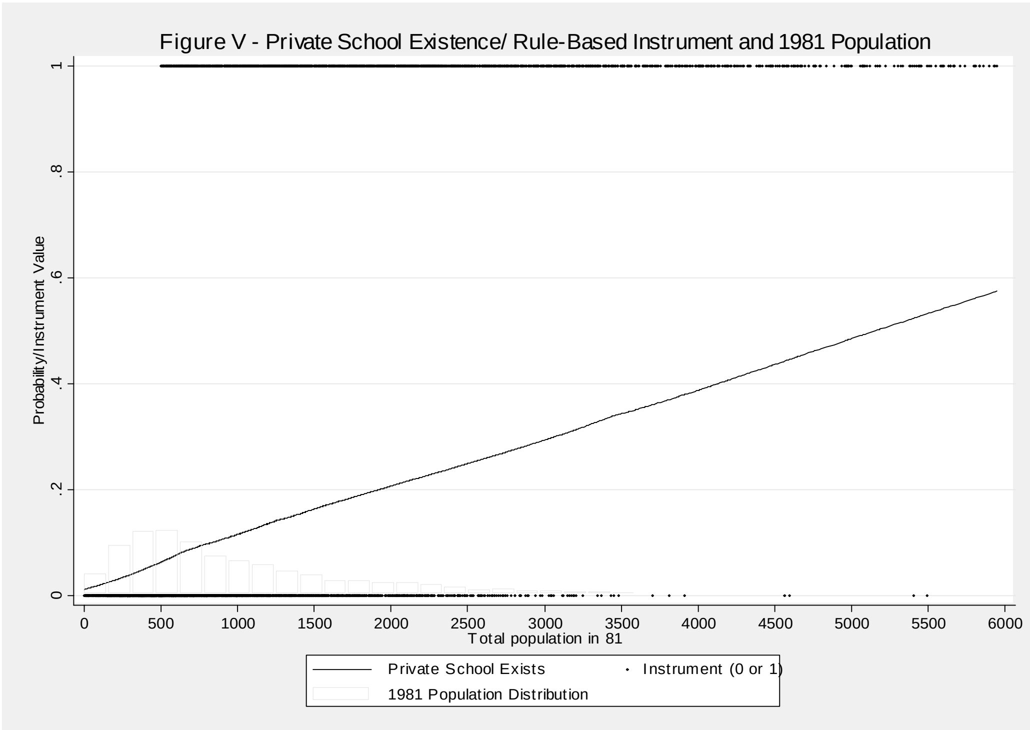
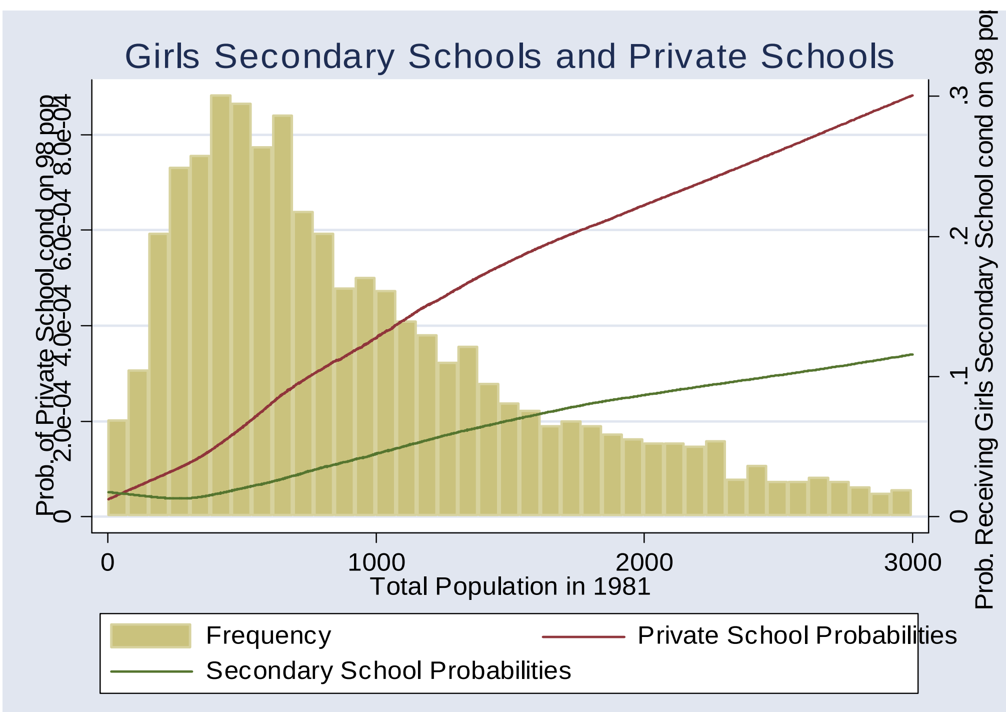


Figure V - Private School Existence/ Rule-Based Instrument and 1981 Population
Probability/Instrument Value
0
.2
.4
.6
.8
1
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
Total population in 81
Private School Exists
Instrument (0 or 1)
1981 Population Distribution

**Fig VI**


Girls Secondary Schools and Private Schools
Prob. of Private School cond on 98 pop
0
2.0e-04
4.0e-04
6.0e-04
8.0e-04
Prob. Receiving Girls Secondary School cond on 98 pop
0
.1
.2
.3
0
1000
2000
3000
Total Population in 1981
Frequency
Private School Probabilities
Secondary School Probabilities